Model Formulation ■

A Self-scaling, Distributed Information Architecture for Public Health, Research, and Clinical Care

Andrew J. McMurry, Clint A. Gilbert, Ben Y. Reis, PhD, Henry C. Chueh, MD, MS, Isaac S. Kohane, MD, PhD, Kenneth D. Mandl, MD, MPH

Abstract **Objective:** This study sought to define a scalable architecture to support the National Health Information Network (NHIN). This architecture must concurrently support a wide range of public health, research, and clinical care activities.

Study Design: The architecture fulfils five desiderata: (1) adopt a distributed approach to data storage to protect privacy, (2) enable strong institutional autonomy to engender participation, (3) provide oversight and transparency to ensure patient trust, (4) allow variable levels of access according to investigator needs and institutional policies, (5) define a self-scaling architecture that encourages voluntary regional collaborations that coalesce to form a nationwide network.

Results: Our model has been validated by a large-scale, multi-institution study involving seven medical centers for cancer research. It is the basis of one of four open architectures developed under funding from the Office of the National Coordinator of Health Information Technology, fulfilling the biosurveillance use case defined by the American Health Information Community. The model supports broad applicability for regional and national clinical information exchanges.

Conclusions: This model shows the feasibility of an architecture wherein the requirements of care providers, investigators, and public health authorities are served by a distributed model that grants autonomy, protects privacy, and promotes participation.

■ **J Am Med Inform Assoc.** 2007;14:527–533. DOI 10.1197/jamia.M2371.

Introduction

We describe our self-scaling, distributed architecture for health data exchange that meets the needs of public health, research, and care delivery. The work reported here builds on the Shared Pathology Informatics Network (SPIN)[1–13] as a model to protect patient privacy, grant institutional autonomy, and exploit legacy systems and data sharing agreements. This approach has been successfully used nationally[14,15] to share Health Insurance Portability and Accountability Act (HIPAA) de-identified[16] human specimens.[17–19] SPIN also influenced key aspects of the Markle Foundation's Connecting for Health Framework. (Shirky C, personal communication, 2005).[20,21] Recognizing its broad applicability for exchanging clinical information, the SPIN model has been extended to satisfy the biosurveillance use case[22] as defined by the American Health Information Community (AHIC). Through these examples, we demonstrate SPIN as a prototype architecture for the National Health Information Network (NHIN).[22–24]

Affiliations of the authors: Children's Hospital Informatics Program at the Harvard–MIT Division of Health Sciences and Technology (AJM, CAG, BYR, ISK, KDM), Dana-Farber/Harvard Cancer Center (AJM), Harvard Medical School (BYR, ISK, KDM), and the Laboratory of Computer Science, Massachusetts General Hospital (HCC), Boston, MA.

Supported by contract N01-LM-3-3515 and grant 1 R01 LM007677-01 from the National Library of Medicine, grant P01 CD000260-01 from the Centers for Disease Control and Prevention, grant 5P30CA06516-40 from the National Cancer Institute, and contract number 5225 3 338CHI from the Massachusetts Department of Public Health.

The authors thank the dedicated members of the AEGIS development team, whose contributions were critical in making this effort a success: Lucy Hadden, Chaim Kirby, Chris Cassa, Karen Olson, and Lucas Jordan. Countless individuals contributed to the overall SPIN mission. Of those not mentioned above, the following people also contributed toward the development of SPIN for public health: Ana Holzbach, David Berkowicz, and Connie Gee.

Correspondence and reprints: Andrew J. McMurry, Children's Hospital Informatics Program at the Harvard–MIT Division of Health Sciences and Technology, 300 Longwood Ave., Enders Room 150, Boston, MA 02115; e-mail: <amcmurry@chip.org>.

Received for review: 1/07/2007; accepted for publication: 4/09/2007.

Background

Significance

Motivated by the need to detect infectious disease outbreaks, track influenza, and provide early warnings of bioterrorism, the AHIC has made biosurveillance a top priority for the NHIN.[22] There is a growing consensus[25] that a successful NHIN must standardize information storage and messaging formats, address privacy concerns, accurately identify patients, and resolve varying local, state, and federal regulations. These issues are pervasive across the NHIN use cases.[22–24] For example, the biosurveillance use case requires both national anonymized coverage for routine analysis and provider authorized re-identification during emergency investigations. Importantly, our approach

not only fulfills the biosurveillance requirements but also supports research and routine clinical care on the same network.

Shared Pathology Informatics Network

Our NHIN architecture extends SPIN,[1–13] which was originally funded[1,26] by the National Cancer Institute to link the vast collections of human specimens that are infrequently shared for cancer research.[18,19] SPIN sets forth the institutional agreements and distributed database architecture to grant institutional autonomy and protect patient privacy according to HIPAA regulations. SPIN has successfully completed a feasibility study involving seven independent medical centers sharing millions of human specimens.[14,15]

Using a peer-to-peer architecture, institutions become SPIN members (nodes) by securing institutional review board (IRB) approvals and deploying the SPIN software. At any time, an institution can withdraw from the network without leaving their data behind or disabling the network. SPIN nodes can serve as "peers" or "supernodes" to query[8] local databases or networks of child nodes, respectively.

SPIN allows institutions to expose[9] de-identified pathology reports while keeping corresponding reports containing Protected Health Information (PHI)[16] disconnected from the Internet. A randomly generated unique identifier is assigned to both the PHI and de-identified reports in a locally controlled "codebook." The machine storing the codebook is disconnected from the Internet and protected according to each participating site's policies. The resulting solution is flexible and compliant with HIPAA regulations.

SPIN provides three levels of increasing access commensurate with investigator credentials and IRB approvals.[13] First, feasibility studies are conducted using a statistical level query that returns only aggregated results. Second, individual de-identified cases are selected by investigators certified by one of the participating institutions. The third level allows requests for specimens and clinical data that must be approved by the institution storing the requested data. Figure 1 illustrates the SPIN software components that enable increasing levels of investigator access.

Biosurveillance Use Case

The AHIC use case[22] calls for a system that can aggregate biosurveillance data from a network of organizations, use existing data-sharing agreements, monitor patient disclosures, credential investigators, and ensure timely adoption. Implicit within these goals are patient de-identification during routine analysis and patient re-identification during emergency investigation. The AHIC use case is not implementation-specific and allows a wide range of transport methods (push and pull), access policies (HIPAA and institutional agreements), and identification systems (for patients and public health investigators). Many of these challenges were already addressed in whole or in part by the SPIN research effort, prompting us to develop an extension of SPIN to support clinical information exchange.

Formulation Process

The design of the SPIN architecture allowed us to adopt the SPIN distributed database, peer group routing subsystem, and query protocol without modification. However, the requirements of the biosurveillance use case are beyond the original intent of the SPIN research network. For instance, real-time surveillance could no longer use the codebook approach because clinical records must be immediately available during public health investigations. Also, we needed to resolve the variations in disclosure policies across institutions, states, and types of investigations (now going beyond research to include public health); disclosures could no longer be authorized by IRB approvals alone.[27] Furthermore, public health investigators are not members of the hospital nodes, posing additional challenges to their identification and credentialing. The need for immediate access and external authority prompted significant enhancement of the SPIN de-identification, authorization, and auditing frameworks. Because these authorizations are commensurate with the strength of evidence of abnormal disease activity, we also needed a means to re-identify only the patients who signaled a potential public health threat. The resulting SPIN-based biosurveillance architecture is illustrated in Figure 2.

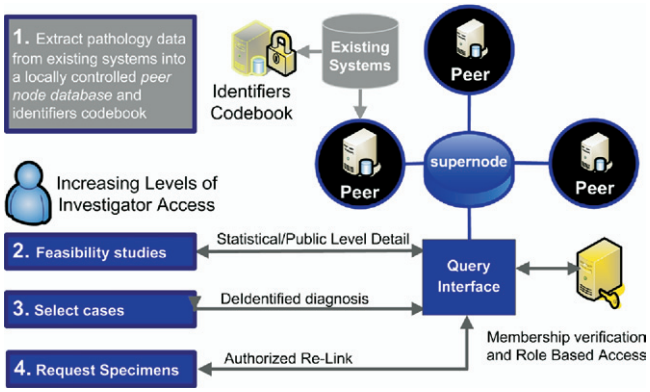

Figure 1. SPIN software components enabling increasing levels of investigator access. (1) Extract pathology data from existing systems into a locally controlled peer database and codebook. (2) Perform feasibility studies using public-level access to population health statistics. (3) Select individual cases of interest after reviewing de-identified patient diagnoses. (4) Request specimens with IRB approval, re-linking cases to codebook entries.

Model Description

Our architecture implements the biosurveillance use case and fulfills five desiderata. First, we adopt a distributed database to prevent creation of a monolithic repository, vulnerable to breach or misuse. Second, we enable strong institutional control to engender participation by the care delivery organizations and laboratories that provide data.[28] Third, we ensure the accountability and oversight necessary to ensure public trust and protect privacy.[29] Fourth, we facilitate real-time analysis of anonymized clinical records. Re-identification occurs only during a public health investigation, using only the cohort of encounters that signaled potential threat. This process happens under institutional control, according to hospital policies and commensurate with the needs of public health and the certified authority of investigators.[30] Fifth, we ensure that our architecture is technically and socially scalable to extend participation, allow new data sources and applications, and facilitate voluntary collaborations in line with the goals of the National Health Information Network.[31]

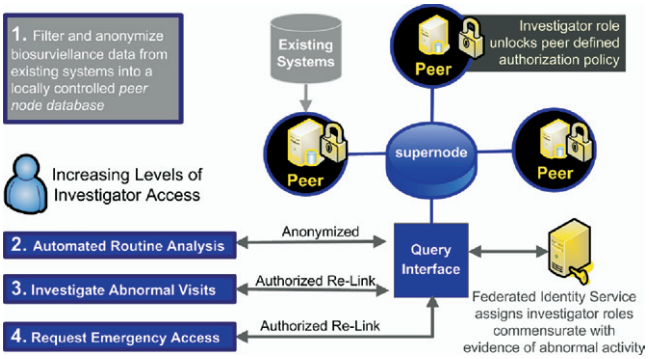

Figure 2. Real-time biosurveillance architecture. (1) Filter, anonymize, format, and load biosurveillance data from existing systems into a locally controlled peer. (2) Perform automated routine analysis to detect unusual patterns of disease; notify public health agencies. (3) Public health agencies investigate abnormal cases, getting more details in real time after presenting proper credentials and being authorized by the care provider. (4) If evidence of abnormal activity is detected, public health investigators request emergency-level access to patient records by presenting proper credentials and being authorized by the care provider.

Distributed Database

To leverage existing hospital databases and legacy information systems, we provide a 3-step pipeline of extraction, transformation, and loading modules (Fig. 3). First, patient records are extracted from local databases or extensible markup language (XML) files. The extracted records are then anonymized, for example by blurring geocoded home addresses, so that patient identity is protected, but sufficient location information is transmitted to detect clustering of cases.[32] Other identifiers, such as patient names found in free text reports,[4] also are removed. Each patient record is assigned a random link identifier to allow re-identification during investigations. Autocoding engines[5,33] then transform free text input into a standard medical vocabulary. Finally, the anonymized and coded data are loaded into the peer database.

Institutions exchange digital certificates with approved peers to certify their identity and secure communications, forming "peer groups" (Fig. 4). Peer groups allow a single institution to concurrently participate in multiple public health, research, and clinical information exchanges. Because fully meshed networks require approval from every other institution, hub and spoke models are more commonly used. Hub and spoke models minimize the number of peer relationships using a single entry point (supernode) for each peer group.

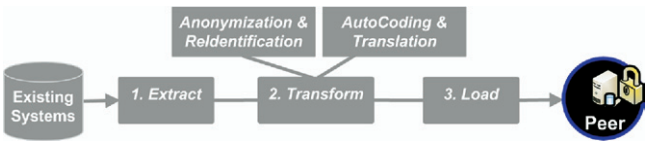

Figure 3. Three-stage pipeline: extract, transform, and load. The modular design allows pluggable transformer definitions to anonymize patients and process free text input.

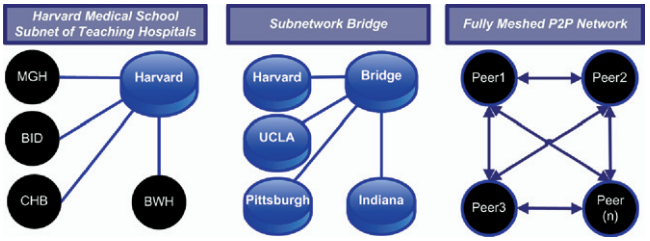

Figure 4. Three examples of peer group configurations. Individual peer institutions are displayed in black, supernodes are displayed in light blue. From left to right: (1) Subnetwork of Harvard Medical School teaching hospitals. (2) National SPIN network bridging subnetworks. (3) Fully meshed peer-to-peer network.

The SPIN-distributed query interface allows all members of a peer group to be contacted with a single query. Queries are performed by contacting the root supernode of the peer group, which propagates the message to each peer network or subnetwork until all peers are contacted. Results are aggregated asynchronously in reverse order. From the perspective of a client using the query interface (Fig. 5), there is no difference between a SPIN network query and a local query.

Institutional Autonomy and Distributed Access Controls

Although secure transmission methods prevent third parties from "listening in," over-the-wire encryption does not ensure that queries are made in good faith. Methods to identify and authorize[34] public health investigators also are needed. This is challenging because peers cannot be expected to host up-to-date registries of investigators' identities, yet hospitals must also remain able to authorize all disclosures. To address this, our framework allows trusted agencies to certify the identities and roles of investigators. Each institution specifies what is allowed to be disclosed for each role according to that institution's policies. This required us to build a Distributed Access Control Framework, as illustrated in Figure 6.

Investigator Accountability and Patient Trust

All investigation scenarios record logging statements at each peer that cannot be removed by external parties.[29] These logs contain the certified identity of the investigator, the identity of the trusted agency who certified investigation, and the time of query. Controversial credentialing of investigators may provide immediate disclosures but will always record an indelible audit trail, preventing clandestine investigations. Care providers are able to challenge the reasonableness of agencies' queries and deny access to agencies if they do not keep patient and provider interests paramount.

Similarly, patients can audit care provider policies and investigator disclosures. This transparency is provided by the SPIN-distributed query, which returns the policies and query logs from all peers. Because all peer group members receive and log the same broadcasted query, a single institution cannot turn off logging or hide disclosures without coming under immediate scrutiny.

Real-Time Anonymized Analysis and Patient Re-identification

SPIN enables increasing levels of investigator access with peer-controlled disclosure. In the research case, investiga-

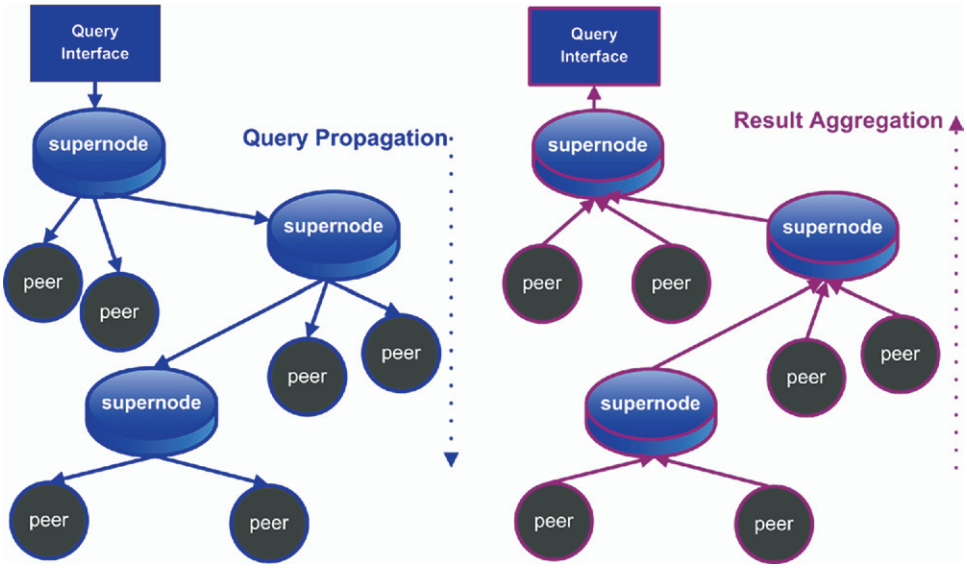

Figure 5. Asynchronous query broadcast and result aggregation on the SPIN network.

tors first review anonymized reports to initiate feasibility studies. Next they select a handful of cases for IRB-authorized research. We will now provide another validation of this principle with respect to public health surveillance. Using the SPIN approach, our Automated Epidemiologic Geotemporal Integrated Surveillance (AEGIS)[35] biosurveillance system provides aberration detection, incurs minimal risk to patient privacy, and allows timely investigations to occur under emergency conditions.

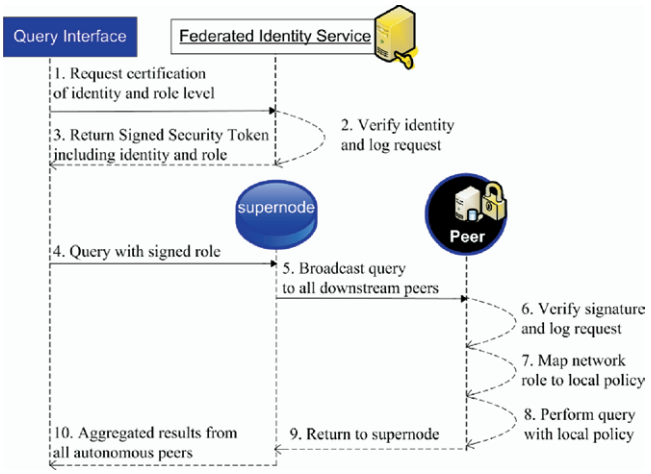

Figure 6. Distributed access control framework. (1) An investigator requests certification of her identity and role from the mutually trusted Federated Identity Service. (2) The Federated Identity Service verifies and logs the certification request. (3) The Federated Identity Service returns a signed security token certifying the investigator's identity and role. (4) The investigator then issues a distributed query, attaching the signed security token. (5) The distributed query is performed as usual; the query and credentials are broadcast to each downstream peer. (6) Each peer verifies the digital signature of the signed security token and checks for expiry. (7) Each peer loads the peer-specific disclosure policy. (8) The policy determines which data elements are returned during the local query. (9) Peers return results to the supernode that issued the distributed query. (10) The supernode returns the aggregated results from all peers.

Modern biosurveillance approaches rely on data mining to search for unusual patterns of disease.[36] Hence, the algorithms require information on all encounters from all care provider locations.[37] This qualitatively and quantitatively shifts the privacy tradeoff balance: disclosure of patient identity is necessary on only the subset of patients that is part of an identified or suspected outbreak, but automated analysis of all patient encounters is necessary to detect clustering.

Most biosurveillance systems look for spatial clustering among cases to signal possible outbreaks. The case locations are often based on patients' home addresses, which are very identifying even if transmitted as geocoded coordinates or plotted on a map.[38,39] Therefore, to preserve patient anonymity, the resolution location data for patients often is reduced to the zip code level. Although sharing patients' zip codes is allowable under the HIPAA limited dataset, the loss in resolution diminishes the effectiveness of cluster detection algorithms.[40] To preserve privacy while retaining cluster-detection power, we have implemented an algorithm that blurs the geocoded coordinates according to the underlying population density.[32] We share the anonymized (blurred) addresses routinely, but share the precise addresses only when a cluster possibly signifying an outbreak is detected and there is agreement between the public health agency and the institutional data source that the data should be shared. SPIN provides a mechanism for increasing levels of investigator access commensurate with public health need and hospital policies (Table 1). First, a routine analysis query returns anonymized details of all patient encounters within a specified time period. When aberrations are detected, the investigator may re-authorize with a higher level of access and re-identify the list of abnormal patient visits.

Routine Analysis

The AEGIS biosurveillance system performs automated real-time analysis of anonymized patient encounters from all

Table 1 ■ Examples of Peer-Specific Authorization Policies With Increasing Levels of Investigator Access

	Routine Analysis	Alarm Investigation	Emergency Investigation
Visit identification	Anonymize	Permit	Permit
Gender	Permit	Permit	Permit
Chief complaint	Permit	Permit	Permit
Location	Anonymize	Anonymize	Permit
Disposition		Permit	Permit
Temperature		Permit	Permit
Check-in time		Permit	Permit
Discharge time		Permit	Permit
MRN			Permit

participating care providers, running geospatial and temporal detection algorithms. Due to seasonal and other trends in the data, the algorithms perform better when given a long historical baseline to compare with current health care activity. SPIN provides broad, regional access (Table 1) to anonymized data while protecting patient identity and reinforcing institutional control.

Alarm Investigation

When aberrations are detected, alarm notifications are sent to public health agencies with an anonymized summary of the patient encounters that prompted investigation. If further investigation is necessary, officials will increase their access level network-wide by certifying the alarm investigation role (Table 1). This role can be used to request more detailed information about the aberrant patient encounters. The hospital will then return more detailed information for only those patients who signaled the alarm. This occurs either in an open loop mode, in which a person at the institution adjudicates each investigation, or in a closed loop mode, in which the institution returns identifying data if its policies allow it and the querying investigator presents an appropriate role and signed security token (Fig. 6).

Emergency Investigation

To ensure a rapid public health response under emergency conditions, we created a permanent closed loop mode for public health authorities in which they can exercise broad investigative powers. Individual institutions still are required to authorize this role, and as with all queries, accountability is enforced post hoc with audit trails. This requires an emergency level role (Table 1) and re-identification similar to that performed in the alarm investigation case. For care providers or institutions uncomfortable with disclosing patient PHI under any circumstance, the local emergency contact information (for example, the infection control nurse) can be provided in lieu of patient records. Investigation may continue through manual lookup of patient records at the source institution using the anonymized link identifiers or medical record numbers.

Self-Scaling Architecture Promoting Timely NHIN Deployment

The idea that the NHIN will be grown from the bottom up and not top down is gaining acceptance.[25] We assert that our model fulfills this self-scaling need in the following respect: this architecture promotes individual participation and collaboration among Regional Health Information Organizations (RHIOs).[41] As shown in Figure 4, a RHIO directly corresponds to a SPIN peer group. Autonomous peers form larger peer groups, and peer groups themselves, can be linked to form larger, networked communities. Autonomy is central to this organizational trust, and ensures that care providers remain stewards of patient privacy.

The SPIN model seeks to expedite early NHIN deployment by leveraging legacy information systems and existing institutional policies. For example, the federated identity and distributed access controls allow hospitals to continue using IRB and HIPAA authorizations. Other examples include the submission tools and query interfaces that extract, transform, and share data from existing databases using standard medical vocabularies. We believe the only way to ensure early participation is to make the technical and procedural burden as light as possible.

Validation Examples

Research

SPIN has demonstrated both national and regional viability for multi-institution cancer research efforts. On a national scale, SPIN investigators have completed a feasibility study involving seven large medical centers sharing a collective library of millions of annotated human specimens.[14,15] On a regional scale, an operational version is deployed and in use at the Dana-Farber Harvard Cancer Center.[3]

Public Health

This model is an essential component of one of four open architectures developed with funding from the Office of the National Coordinator of Health Information Technology, and this model fulfills the AHIC biosurveillance use case.[22] In January 2007, this architecture was presented to the AHIC stakeholders using the live AEGIS system developed for the Massachusetts Department of Public Health.

Clinical Care

Clinical applications within the NHIN will require complete patient histories to be available regardless of where a patient receives care. Using the SPIN query interface, patients and physicians could locate records distributed across the network. Queries across the system could return data populating electronic health records or personally controlled health records.[45,47,48] Patients also could authorize disclosures, review HIPAA compliant audit trails, and even consent to research for which they stand to benefit.[49]

Discussion

Significance

Many scientists are calling for a closer connection between translational research and routine patient care. Although human specimens and patient histories represent a valuable resource in the postgenomic era, few investigators have authorization across all locations where patients receive care.[27]

We developed SPIN to link existing databases while building institutional agreement and protecting patient privacy. As a result, SPIN has been deployed across multiple locations and has addressed pervasive issues in sharing patient data. The broad applicability of this approach allowed us to develop a public health infrastructure with minimal effort. Specifically, we leveraged the distributed database to survey health statistics and detect disease aberrations. SPIN also

Table 2 ■ A Single Peer-to-Peer Network to Support Public Health, Research, and Clinical Care

	Statistical	Nonidentifying	Identified Dataset
Research	Feasibility studies	Case selection	Institutional review board approved research
Public health	Health statistics	Aberration detection	Emergency investigation
Care delivery	Environmental factors	Quality improvements	Informed clinical care

provided patient re-identification capabilities during public health emergencies, as well as the audit trails and authorizations required for public health investigations.

With a fully deployed SPIN infrastructure, it would be possible to simultaneously support public health, research, and clinical care activities (Table 2) on a single peer-to-peer network. New participants and applications can extend this self-scaling architecture without interrupting critical services or participation agreements. This loosely coupled, bottom-up approach ensures that a SPIN-style NHIN would be scalable and broadly deployable.

Limitations and Future Work

The architecture that we have described has technical limitations that we will address in future work. Efforts are underway to use public-key cryptography to encrypt peer databases and store the decrypting keys remotely. With encrypted storage, the data are protected even if the server is physically stolen. We also are working toward the WS-Security[42] standard to prevent intermediate supernodes from inspecting the aggregated responses that pass through them (Figs. 5 and 6).

The tree shape of our proposed network makes queries easy, but means that a failure at a supernode can render child nodes unreachable. Using redundant peers for fault tolerance is possible but requires additional hardware and software complexity. We are in the process of developing a network discovery protocol using the existing distributed query capabilities, allowing the SPIN network to route around offline supernodes.[43,44]

Currently, the distributed biosurveillance network performs anonymized routine analysis at a single location. Thus, when an alarm is generated all participating institutions must trust that requests for identifying patient information are justified by evidence of abnormal activity. We are investigating a distributed computational approach in which all institutions participate in verifying the evidence of aberration, thereby ensuring that all requests for increased access are made in good faith.[29]

Spatial anonymization reduces the sensitivity and specificity of spatial clustering algorithms. However, for the algorithms used by AEGIS, these reductions are small enough to make the privacy gains provided by spatial anonymization worth the cost. We continue to explore new methods to protect privacy and preserve cluster detection effectiveness.

The design of our architecture currently offers no robust mechanism for identifying a single patient across multiple locations. We will continue to explore development of a Record Locator Service,[20] which takes patient identifiers as input and returns either the anonymized record location or a list of records aggregated[21] from all locations where the patient has received care.

Conclusions

Development of an NHIN requires broad participation using the systems and policies already in place. Concerns about patient privacy and institutional control of data are pervasive throughout public health, research, and clinical care. We propose a distributed architecture that grants autonomy, protects privacy, and promotes participation.[46,47]

References ■

1. NIH. Shared Pathology Informatics Network. [RFA] 2000. Available at: http://grants.nih.gov/grants/guide/rfa-files/rfa-ca-01-006.html. Accessed November 8, 2006.
2. Shared Pathology Informatics Network (SPIN). 2006. Available at: http://spin.nci.nih.gov/. Accessed December 8, 2006.
3. Dana-Farber/Harvard Cancer Center: Virtual Specimen Locator (VSL). Available at: http://www.dfhcc.harvard.edu/center-initiatives/strategic-initiatives/virtual-specimen-locator-vsl/. Accessed November 8, 2006.
4. Beckwith BA, Mahaadevan R, Balis UJ, Kuo F. Development and evaluation of an open source software tool for deidentification of pathology reports. BMC Med Inform Decis Mak 2006;6:12.
5. Berman JJ. Automatic extraction of candidate nomenclature terms using the doublet method. BMC Med Inform Decis Mak 2005;5:35.
6. Berman JJ. Pathology data integration with eXtensible Markup Language. Hum Pathol 2005;36:139–45.
7. Gupta D, Saul M, Gilbertson J. Evaluation of a deidentification (De-Id) software engine to share pathology reports and clinical documents for research. Am J Clin Pathol 2004;121:176–86.
8. Holzbach AM CH, Porter AJ, Kohane IS, Berkowicz D. A query engine for distributed medical databases. Medinfo 2004;21:1519.
9. Namini AH, Berkowicz DA, Kohane IS, Chueh H. A submission model for use in the indexing, searching, and retrieval of distributed pathology case and tissue specimens. Medinfo 2004; 11:1264–7.
10. McDonald CJ, Dexter P, Schadow G, et al. SPIN query tools for de-identified research on a humongous database. AMIA Annu Symp Proc 2005:515–9.
11. Thomas SM, Mamlin B, Schadow G, McDonald C. A successful technique for removing names in pathology reports using an augmented search and replace method. Proc AMIA Annu Symp 2002:777–81.
12. Kowalczyk L. Harvard hopes database will speed cancer cures. 2005. Available at: http://www.boston.com/news/globe/health_science/articles/2005/11/21/harvard_hopes_database_will_speed_cancer_cures/. Accessed November 21, 2006.
13. Virtual Specimen Locator: User Guide. Available at: https://querytool.med.harvard.edu/querybuilder/pdf/VSL-User-Guide.pdf. Accessed December 6, 2006.
14. Drake TA, Braun J, Marchevsky A, et al. A system for sharing routine surgical pathology specimens across institutions: the Shared Pathology Informatics Network (SPIN). Hum Pathol (e-pub ahead of print) May 8, 2007. DOI 10.1016/j.humpath.2007.01.007.
15. Patel AA, Gupta D, Seligson D, et al. Availability and quality of paraffin blocks identified in pathology archives: a multi-institu-

tional study by the Shared Pathology Informatics Network (SPIN). BMC Cancer 2007;7:37.
16. Standards for privacy of individually identifiable health information: final rule. Fed Regist 2002;67:53181–273.
17. Protection of human subjects: common rule. Fed Regist 1991;56: 28003–32.
18. The National Biospecimen Network (NBN) blueprint. Available at: http://biospecimens.cancer.gov/nbn/blueprint.asp. Accessed November 8, 2006.
19. Eiseman E, Bloom G, Brower J, Clancy N, Olmsted SS. Case Studies of Existing Human Tissue Repositories. Santa Monica: RAND, 2003.
20. The Connecting for Health Common Framework. Available at: http://www.connectingforhealth.org/. Accessed December 5, 2006.
21. The Common Framework: Technical Issues and Requirements for Implementation. Available at: http://www.connectingforhealth.org/commonframework/docs/T1_TechIssues.pdf. Accessed June 12, 2007.
22. Office of the National Coordinator for Health Information Technology O. Harmonized Use Case for Biosurveillance (Visit, Utilization and Lab Result Data). 2006. Available at: http://www.hhs.gov/healthit/documents/BiosurveillanceUseCase.pdf. Accessed December 1, 2006.
23. Office of the National Coordinator for Health Information Technology O. Harmonized Use Case for Electronic Health Records (Laboratory Result Reporting). 2006 March 19, 2006 [cited December 8, 2006]. Available at: http://www.hhs.gov/healthit/documents/EHRLabUseCase.pdf. Accessed December 1, 2006.
24. Office of the National Coordinator for Health Information Technology O. Harmonized Use Case for Consumer Empowerment (Registration and Medication History). 2006. Available at: http://www.hhs.gov/healthit/documents/ConsumerEmpowerRegMedUseCase.pdf. March 19, 2006.
25. Office of the National Coordinator for Health Information Technology. Summary of Nationwide Health Information Network (NHIN) Request for Information (RFI) Responses. 2005 [cited December 8, 2006]; Available at http://www.hhs.gov/healthit/rfisummaryreport.pdf. Accessed December 1, 2006.
26. Kohane IS. Consented High-Performance Indexing and Retrieval of Pathology Specimens (CHIRPS). 2000. Available at: http://spin.nci.nih.gov/CHIRPSGrant.pdf. Accessed November 6, 2006.
27. McWilliams R, Hoover-Fong J, Hamosh A, Beck S, Beaty T, Cutting G. Problematic variation in local institutional review of a multicenter genetic epidemiology study. JAMA 2003;290: 360–6.
28. Sittig DF, Shiffman RN, Leonard K, et al. A draft framework for measuring progress towards the development of a National Health Information Infrastructure. BMC Med Inform Decis Mak 2005;5:14.
29. Weitzner DJ AH, Berners-Lee T, Hanson C, et al. Transparent Accountable Data Mining: New Strategies for Privacy Protection. Computer Science and Artificial Intelligence Laboratory Technical Report. January 27, 2006. Available at: http://hdl.handle.net/1721.1/30972. Accessed November 1, 2006.
30. Lazarus R, Yih K, Platt R. Distributed data processing for public health surveillance. BMC Public Health 2006;6:235.
31. Overhage JM, Evans L, Marchibroda J. Communities' readiness for health information exchange: The national landscape in 2004. J Am Med Inform Assoc 2005;12:107–12.
32. Cassa CA, Grannis SJ, Overhage JM, Mandl KD. A context-sensitive approach to anonymizing spatial surveillance data: impact on outbreak detection. J Am Med Inform Assoc 2006;13: 160–5.
33. Tsui FC, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. J Am Med Inform Assoc 2003;10:399–408.
34. caBIG Security Technology Evaluation—White Paper. 2006 [cited December 8, 2006]; Available at: https://cabig.nci.nih.gov/workspaces/Architecture/Documents/Arch_Workspace/caBIG_Technology_Evaluation_Security_White_Paper_version_0_2.pdf. Accessed December 1, 2006.
35. Reis BY, Kirby C, Sprecher E, et al. Advanced Modular Design for Scalable Biosurveillance Systems. Advances in Disease Surveillance 2006:1.
36. Mandl KD, Overhage JM, Wagner MM, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11:141–50.
37. Reis BY, Mandl KD. Integrating syndromic surveillance data across multiple locations: effects on outbreak detection performance. AMIA Annu Symp Proc 2003; 549–53.
38. Brownstein JS, Cassa CA, Mandl KD. No place to hide—reverse identification of patients from published maps. N Engl J Med 2006;355:1741–2.
39. Brownstein JS, Cassa CA, Kohane IS, Mandl KD. Reverse geocoding: concerns about patient confidentiality in the display of geospatial health data. AMIA Annu Symp Proc 2005:905.
40. Olson KL, Grannis SJ, Mandl KD. Privacy protection versus cluster detection in spatial epidemiology. Am J Public Health 2006;96:2002–8.
41. Deasy B. Electronic Health Record Interchange. A Response to the Federal Government's Request for Information on the Development and Adoption of a National Health Information Network. 2005. Available at: http://www.thecre.com/pdf/CapTechNHINComments.pdf. Accessed December 8, 2006.
42. Nadalin A, Kaler K, Monzillo R, Hallam-Baker P. Web Services Security: SOAP Message Security 1.1 (WS-Security 2004). 2006. Available at: http://docs.oasis-open.org/wss/v1.1/. Accessed December 8, 2006.
43. Alexandru I, Pawel G, Johan P, Dick E. Correlating Topology and Path Characteristics of Overlay Networks and the Internet. Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid (CCGRID'06). IEEE Computer Society 2006;2:10–16.
44. Praveen K, Sridhar G, Sridhar V. Bandwidth and Latency Model for DHT Based Peer-to-Peer Networks under Variable Churn. Proceedings of the 2005 Systems Communications (ICW'05, ICHSN'05, ICMCS'05, SENET'05). IEEE Computer Society 2005:00.
45. Simons WW, Mandl KD, Kohane IS. The PING personally controlled electronic medical record system: technical architecture. J Am Med Inform Assoc 2005;12:47–54.
46. Adida B, Kohane IS. GenePING: Secure, scalable management of personal genomic data. BMC Genomics 2006;7:93.
47. Mandl KD, Szolovits P, Kohane IS. Public standards and patients' control: how to keep electronic medical records accessible but private. BMJ 2001;322:283–7.
48. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. 2006;13:121–6. Epub Dec 15, 2005.
49. Kohane IS, Altman RB. Health-information altruists—a potentially critical resource. N Engl J Med 2005;353:2074–7.